

# The freshness of Persian information: the case of Persian newspapers indexed in Google, Yahoo and Bing

M. Yaghtin<sup>1</sup> | Z. Honarjooyan<sup>2</sup> | H. Sotudeh<sup>3</sup>

1. PhD student of Shiraz University      yaghtin.maryam@gmail.com
2. [Corresponding Author] MA in Knowledge and Information  
Sciences      z.honarjooyan@gmail.com
3. Faculty Member of Shiraz University      sotudeh@shirazu.ac.ir

Iranian Journal of  
**Information  
Processing &  
Management**

Iranian Research Institute

for Information Science and Technology  
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 30 | No. 3 | pp. 737-758

Spring 2015

<https://doi.org/10.35050/JIPM010.2015.031>



**Abstract:** Aiming to study the freshness of Persian information, this study attempted to explore the indexing speed of Persian newspapers in Google, Yahoo and Bing. The present study was conducted using a webometric method. The population of the study consisted of all online newspapers published in Persian. The study sample was identified using a targeted method. It included those Persian newspapers listed on the MagIran database with their electronic versions regularly published on their respective websites. The results revealed that Google outperforms the two other search engines in terms of timely indexing of the Persian newspapers. Furthermore, it generally enjoys a more consistent and stable indexing model. Yahoo and Bing are not only slower in indexing the newspapers, but also experience a comparably lower consistency level in their indexing models. Yahoo and Bing show to be comparable in their indexing speeds. In this study, the indexing status of Persian newspapers on different internet search engines was investigated for the first time. Results showed that those accessing these newspapers using internet search engines will have a higher chance of accessing their most updated versions using Google.

**Keywords:** Persian Newspapers; Google; Yahoo; Bing; Indexing Speed; Freshness Rate; Immediate Indexing

# بررسی تازگی اطلاعات فارسی: مطالعه موردی روزنامه‌های فارسی نمایه‌شده در پایگاه‌های گوگل، یاهو و بینگ

مریم یقظین<sup>۱</sup> | زهره هنرجویان<sup>۲</sup> | هاجر ستوده<sup>۳</sup>

۱. [پدیدآور رابط] دانشجوی دکتری علم اطلاعات و دانش‌شناسی؛ دانشگاه شیراز  
yaghtin.maryam@gmail.com
۲. دانش‌آموخته کارشناسی ارشد علم اطلاعات و دانش‌شناسی؛ کارشناس ارشد کتابخانه شهید دستغیب؛ دانشگاه شیراز  
z.honarjooyan@gmail.com
۳. دکتری علم اطلاعات و دانش‌شناسی؛ استادیار؛ بخش علم اطلاعات و دانش‌شناسی؛ دانشگاه شیراز  
sotudeh@shirazu.ac.ir

## مقاله پژوهشی

دریافت: ۱۳۹۳/۰۵/۲۱

پذیرش: ۱۳۹۳/۰۸/۲۵

## دانشگاه بهداشت اطلاعات

فصلنامه | علمی پژوهشی

پژوهشگاه علوم و فناوری اطلاعات ایران  
(ایرانداک)

شاپا (چاپی) ۸۲۳-۲۵۱

شاپا (الکترونیکی) ۸۳۱-۲۵۱

نمایه در SCOPUS، ISI، و LISTA

jipm.irandoc.ac.ir

دوره ۳۰ | شماره ۳ | صص ۷۳۷-۷۵۸

بهار ۱۳۹۴

<https://doi.org/10.35050/IJPM010.2015.031>

**چکیده:** پژوهش حاضر که با هدف بررسی تازگی اطلاعات فارسی انجام شده، می‌کوشد سرعت نمایه‌سازی روزنامه‌های فارسی را در سه موتور کاوش گوگل، یاهو و بینگ مطالعه و مقایسه کند. این پژوهش به‌روش وب‌سنجی انجام شده و جامعه آن را کلیه روزنامه‌های فارسی پیوسته تشکیل می‌دهد. نمونه پژوهش به‌روش هدفمند شناسایی شد و شامل آن دسته از روزنامه‌های فارسی‌زبان بود که فهرست آنها در پایگاه «مگ‌ایران» آمده و نسخه الکترونیکی آنها به‌طور منظم بر وب‌سایت مربوطه منتشر می‌شود. نتایج حاصل از پژوهش نشان داد که موتور کاوش گوگل از لحاظ سرعت روزآمدسازی روزنامه‌های فارسی، رتبه بالاتری نسبت به موتورهای کاوش یاهو و بینگ دارد. علاوه بر این، گوگل به‌طور کلی از الگوی روزآمدسازی منظم‌تری پیروی می‌کند. دو موتور کاوش یاهو و بینگ نه‌تنها از سرعت کمتری نسبت به گوگل برخوردارند، بلکه الگوی روزآمدسازی نامنظم‌تری نیز دارند. دو موتور کاوش یاهو و بینگ از لحاظ سرعت روزآمدسازی تفاوتی با یکدیگر ندارند. این پژوهش برای نخستین بار وضعیت نمایه‌سازی روزنامه‌های فارسی‌زبان را در موتورهای کاوش وب مورد بررسی قرار داد و نشان داد که جستجوگرانی که با کمک موتورهای کاوش به این روزنامه‌ها دست می‌یابند، در گوگل از بخت بالاتری برای دستیابی به نسخه روزآمد این روزنامه‌ها برخوردارند.



**کلیدواژه‌ها:** روزنامه فارسی؛ گوگل؛ یاهو؛ بینگ؛ نرخ تازگی؛ سرعت روزآمدسازی؛ روزآمدسازی بی‌درنگ

## ۱. مقدمه

وب، با سرشت پویا و رشد روزافزون خود پدیده اطلاعاتی متفاوتی را به ارمغان آورده است. منابع وبی با سرعتی چشمگیر افزایش می‌یابند یا دگرگون می‌شوند. در هر هفته حدود ۳۲۰ میلیون صفحه جدید در وب به وجود می‌آید. حدود ۲۰ درصد از صفحات موجود امروز ظرف یک سال از بین می‌روند و حدود ۵۰ درصد از کل محتوای صفحات تغییر می‌کند (Ntoulas, Cho, & Olston 2004). این پویایی، نگهداری نمایه‌ای روزآمد را که جستجوی اثربخش در وب را تضمین کند، به چالشی بزرگ بدل کرده است (پاپاپترو، پاپاستاورو، و ساماراس ۱۳۸۴).

نرم‌افزارهای خزنده<sup>۱</sup>، ناکام در هم‌گامی با رشد فزاینده وب، از شناسایی صفحات تازه باز می‌مانند و در نتیجه، بخش تازه‌تر وب همواره از حوزه جستجوی موتورهای کاوش دور می‌ماند. بنابراین، همواره فاصله زمانی زیادی از چند روز تا چند هفته بین اطلاعات تازه و زنده موجود در وب و آنچه موتور کاوش جستجو می‌کند، وجود خواهد داشت. آشکار است که در فاصله زمانی بین شناسایی یک صفحه وب تازه آفریده (یا تغییر یافته) و ثبت آن در موتور کاوش، بازیابی اطلاعات این صفحات (و یا تغییرات آنها) امکان‌پذیر نیست (داوودیان ۱۳۸۰).

از آنجا که دسترسی به اطلاعات روزآمد از اهمیت بسیاری برخوردار است، آگاهی نسبت به این مسئله می‌تواند به جستجوی کارآمد و روزآمد منابع فارسی در موتورهای کاوش کمک کند. با توجه به اقبال عمومی کاربران فارسی‌زبان به استفاده از موتورهای کاوش بین‌المللی، مهاجرتی زبان فارسی در مقایسه با زبان‌هایی چون انگلیسی، پایین بودن نسبی شمار گویش‌وران فارسی در سراسر جهان و کمبود موتورهای کاوش خاص زبان فارسی، پژوهش در این باره اهمیتی ویژه می‌یابد.

از این رو، پژوهش حاضر می‌کوشد سرعت نمایه‌شدن روزنامه‌های فارسی را در سه

1. crawler

موتور کاوش گوگل، یاهو و بینگ بررسی و مقایسه کند تا از این رهگذر ضمن روشن ساختن وضعیت روزآمدی روزنامه‌های فارسی‌زبان در پایگاه‌های موتورهای کاوش، مشخص گردد که کدام موتور کاوش برای دسترسی به آخرین مطالب منتشرشده در این روزنامه‌ها مناسب‌تر است. به این منظور، در این پژوهش کلیه روزنامه‌های فارسی پیوسته<sup>۱</sup> که نسخه الکترونیکی آنها به‌طور منظم در وب‌سایت مربوط منتشر می‌شود، شناسایی شده و سرعت روزآمدسازی آنها در این سه موتور بررسی و مقایسه می‌گردد.

## ۲. پیشینه پژوهش

تازگی<sup>۲</sup> پایگاه اطلاعاتی نمایه‌شده به یکی از عوامل اساسی در تعیین کیفیت نمایه موتور کاوش بدل شده و توجه پژوهشگران بسیاری را به خود معطوف داشته است. از جمله، لواندوسکی و همکارانش با بررسی تازگی وب‌سایت‌های آلمانی در پایگاه‌های موتورهای کاوش وب نشان دادند که گوگل با داشتن بیشترین صفحاتی که به‌طور روزانه روزآمد می‌شوند، بهتر از یاهو و ایم‌اس‌ان عمل می‌کند. ایم‌اس‌ان الگوهای نمایه‌سازی روشنی عرضه کرده، اما فرایند روزآمدسازی یاهو بی‌نظم است. به این ترتیب، با توجه به تفاوت کیفیت نمایه‌های موتورهای کاوش، در هنگام جستجوی مطالب جاری نباید به یک موتور کاوش بسنده کرد (Lewandowski, et al. 2005).

نتایج پژوهش مکیل و همکارانش پیرامون سرعت نمایه‌سازی اخبار خبرگزاری رویترز حاکی از آن است که گوگل، آلتاویستا، و یاهو ظرف کمتر از یک ساعت نسبت به هر رویداد خبری واکنش نشان می‌دهند. دیگر موتورهای کاوش، به جز پورتال آلمانی تی-آن‌لاین<sup>۳</sup> مقالات را پس از گذشت سه ساعت از قرارگرفتن بر سایت خبرگزاری ارائه کردند (Machill 2005).

لواندوسکی دریافت که سرعت روزآمدسازی نمایه گوگل بیش از یاهو و تئوما<sup>۴</sup> است (Lewandowski 2004). نوتس در چند پژوهش، تازگی هشت موتور کاوش را مورد بررسی قرار داد و دریافت که موتورهای کاوش بزرگی چون ایم‌اس‌ان، هات‌بات<sup>۵</sup>، گوگل،

1. online  
2. freshness  
3. T-online  
4. Teoma  
5. HotBot

آل د وب<sup>۱</sup> و آلتاویستا همگی در پایگاه‌های اطلاعاتی خود، صفحاتی روزآمد یا مربوط به یک روز قبل دارند، اما صفحاتی در پایگاه‌های موتورهای کوچکی چون گیگابلاست<sup>۲</sup>، تئوما و وایزنات<sup>۳</sup> یافت می‌شود که تاریخ آنها دست کم به ۴۰ روز پیش باز می‌گردد (Notess 2001-2003). به‌طور کلی، مرور یافته‌های پیشین نشان از آن دارد که سرعت روزآمدسازی نمایه‌های موتورهای کاوش گوناگون متفاوت بوده و در این میان گوگل از وضعیت بهتری برخوردار است.

در ایران کمتر پژوهشی پیرامون تازگی وب‌سایت‌های فارسی در پایگاه‌های موتورهای کاوش صورت گرفته است. پژوهش شریف از محدود تحقیقاتی است که به کیفیت نمایه‌های موتور کاوش پرداخته است. وی با استفاده از روش تجربی، تغییرات عملکرد یاهو و گوگل را از نظر پوشش کمی-زمانی نمایه‌سازی و نقش عناصر فراداده‌ای در رتبه‌بندی صفحه‌های وب بررسی کرد و نشان داد که این دو موتور کاوش در بُعد کمی-زمانی و بُعد کیفی عناصر ابرداده‌ای تفاوت چندانی در عملکرد خود نداشته‌اند (۱۳۹۱). به این ترتیب، کمتر اطلاعاتی درباره‌ی روزآمدی منابع فارسی در پایگاه‌های موتورهای کاوش در اختیار ماست.

### ۳. سؤالات پژوهش

۱. کمینه و بیشینه سرعت روزآمدسازی روزنامه‌های فارسی در پایگاه هر یک از موتورهای کاوش چند روز است؟
۲. چه درصدی از روزنامه‌های فارسی بی‌درنگ در هر موتور کاوش روزآمد شده‌اند؟
۳. آیا اختلاف معناداری بین موتورهای کاوش به‌لحاظ میانگین سرعت روزآمدسازی روزنامه‌های فارسی وجود دارد؟
۴. آیا اختلاف معناداری بین روزنامه‌های فارسی در هر موتور کاوش به‌لحاظ میانگین سرعت روزآمدسازی وجود دارد؟
۵. الگوهای روزآمدسازی روزنامه‌های فارسی در هر موتور کاوش چگونه است؟

---

1. All the Web  
2. Gigablast  
3. Wisenut

#### ۴. روش پژوهش

پژوهش حاضر به روش وب‌سنجی انجام شده است. ابزار گردآوری داده‌ها سیاهه واریسی متشکل از نام روزنامه مورد بررسی، تاریخ نسخه زنده و تاریخ نسخه پنهان بود. منظور از نسخه زنده، نسخه اصلی و جاری صفحه وب روزنامه است که از طریق نشانی آن در اینترنت قابل دسترسی است. تاریخ انتشار شماره‌ای که بر این صفحه قرار دارد، به عنوان تاریخ صفحه زنده ذخیره شد. منظور از نسخه پنهان، «تصویر لحظه‌ای<sup>۱</sup>» یا نسخه‌ای از یک صفحه وب است که به عنوان نسخه پشتیبان در زمانی خاص بر یک سرورس دهنده وب ذخیره می‌شود. در این پژوهش، منظور از نسخه پنهان صفحه‌ای از روزنامه است که در پایگاه موتور کاوش ذخیره شده است و در تاریخ جستجو به دست می‌آید. تاریخ این شماره به عنوان تاریخ نسخه پنهان ذخیره شد. بسته به سرعت روزآمدسازی، شماره روزنامه در پایگاه موتور کاوش می‌تواند همان شماره جاری در نسخه زنده یا شماره‌های گذشته باشد. داده‌ها، برای هر موتور کاوش به‌طور جداگانه و با مراجعه مستقیم به وبسایت‌های روزنامه‌ها و همچنین، مراجعه به پایگاه موتورهای کاوش مورد بررسی استخراج شد و سپس به کمک بسته نرم‌افزار آماری اس‌پی‌اس‌اس و با استفاده از فنون آمار توصیفی (فراوانی، درصد و میانگین) و استنباطی (آزمون همبستگی پیرسون و تحلیل رگرسیون) تجزیه و تحلیل شد.

#### ۴-۱. جامعه و نمونه پژوهش

جامعه پژوهش را کلیه روزنامه‌های فارسی‌زبان که نسخه پیوسته آنها در وب منتشر می‌شود، تشکیل می‌دهد. نمونه مورد پژوهش به روش نمونه‌گیری هدفمند انتخاب شد. نمونه پژوهش را روزنامه‌های فارسی‌زبان دارای نسخه پیوسته وبی که در پایگاه اطلاعات نشریات کشور (مگیران)<sup>۲</sup> پوشش داده شده‌اند، تشکیل داد. در زمان گردآوری داده‌ها، این پایگاه تنها مرجعی بود که فهرستی از روزنامه‌های پیوسته را به همراه نشانی الکترونیکی آنها ارائه می‌داد.

1. snapshot  
2. www.magiran.com

#### ۴-۱-۱. شناسایی روزنامه‌های فارسی

در نخستین گام، ۴۴ روزنامه فارسی پیوسته در تاریخ ۲۸ تیر ۱۳۹۱ به کمک مگیران شناسایی شد. به منظور بررسی روزآمدسازی بر پایه الگویی روزانه، ضروری بود این سیاهه به روزنامه‌هایی محدود گردد که نظم روزانه را در انتشار نسخه پیوسته خود بر روی وب رعایت می‌کنند. همچنین، تاریخ انتشار به‌عنوان تاریخ آخرین روزآمدسازی محتوا مد نظر قرار گرفت.

بررسی وبسایت‌های روزنامه‌های شناسایی شده نشان داد که همه آنها تاریخ انتشار را در محتوای صفحه وب خود درج می‌کنند. امکان مشاهده محتوای هشت وبسایت از وبسایت‌های روزنامه‌ها به دلیل بروز خطا<sup>۱</sup> میسر نشد. همچنین، در مدت زمان گردآوری داده‌ها شش وبسایت روزنامه دست کم توسط یکی از موتورهای کاوش مورد بررسی پوشش داده نشده بودند. مجوز پنج روزنامه نیز لغو شده بود. با حذف این تعداد روزنامه، در پایان، تعداد روزنامه‌های فارسی پیوسته قابل بررسی به ۲۵ عنوان رسید. این عناوین شامل روزنامه‌های آفرینش، ابتکار، ابرار، ابرار اقتصادی، ابرار ورزشی، اعتدال، افکار، ایران، تفاهم، جام جم، جهان صنعت، جوان، حمایت، خبر، دنیای اقتصاد، رسالت، سیاست روز، عصر اقتصادی، عصر آزادی، قدس، کیهان، مردم‌سالاری، نسل فردا، هدف اقتصاد، و وطن امروز است.

#### ۴-۱-۲. انتخاب موتور کاوش

از میان موتورهای کاوش، گوگل، یاهو، و بینگ مورد بررسی قرار گرفتند. دلایل انتخاب این سه موتور در وهله اول پشتیبانی آنها از زبان فارسی بوده است. همچنین، این سه در سال ۲۰۱۲ در میان موتورهای کاوش محبوب در صدر قرار داشتند<sup>۲</sup>. فرض پژوهش حاضر این است که این گرایش جهانی برای ایران نیز صادق است.

#### ۴-۲. روش اجرا

در گام نخست، نشانی وبی<sup>۳</sup> کامل صفحه وب روزنامه به‌طور مستقیم در مکان‌یاب

۱. پیوندهای کور و غیرفعال مانند "The site is no longer available"، "Cannot find server"، "Http 404 not found"

۲. [www.alexa.com](http://www.alexa.com), [www.ebizmba.com](http://www.ebizmba.com), [www.listofsearchengines.org](http://www.listofsearchengines.org)

۳. URL

مرورگر وب جستجو شد تا تاریخ محتوای زنده موجود بر وبسایت به دست آید. لازم به ذکر است که به دلیل انتشار روزانه، معمولاً تاریخ جستجو با تاریخ محتوای زنده موجود بر وبسایت یکسان است؛ مگر در برخی موارد که نسخه روز روزنامه به دلایلی خاص منتشر نشده باشد یا نسخه پیوسته آن بر وب قرار نگرفته باشد.

در گام بعد، نشانی وبی کامل صفحه وب روزنامه در کادر جستجوی موتور کاوش وارد شد. سپس، در رکورد بازبایی شده پیوند "cached" در گوگل و یاهو و پیوند "cached" در page در بینگ دنبال شد. آشکار است که این پیوند در واقع به صفحه وبی زنده روزنامه متصل نیست، بلکه ما را به نسخه جاری صفحه وب در نمایه داخلی پایگاه هدایت می‌کند. گام بعد، استخراج تاریخ انتشار مندرج در محتوای هر صفحه وب حافظه پنهان (تاریخ انتشار نسخه جاری صفحه وب روزنامه در نمایه داخلی پایگاه) بود. نرخ تازگی روزانه بر پایه مقایسه تاریخ محتوای زنده و تاریخ مندرج در نسخه پنهان محاسبه شد. برای مثال، مقدار تازگی روزانه یک نسخه حافظه پنهان مربوط به تاریخ ۸ مهر که در تاریخ ۱۰ مهر ماه جستجو شده، ۲ روز است. بررسی برای هر روزنامه در هر موتور کاوش در ۳۰ روز متوالی (از ۹ مرداد تا ۸ شهریور ۱۳۹۱) به طول انجامید. به این ترتیب، ۲۲۵۰ نرخ تازگی روزانه (۲۵ وبسایت  $\times$  ۳ موتور کاوش  $\times$  ۳۰ روز) به دست آمد. نرخ تازگی هر روزنامه در هر موتور کاوش، بر پایه میانگین نرخ‌های تازگی روزانه تمام صفحات آن محاسبه شد.

حساسیت دسترسی به خبرهای روز، نمایه‌سازی بی‌درنگ را برای روزنامه‌ها ضروری می‌سازد. سرعت روزآمدسازی بی‌درنگ در این پژوهش از صفر تا یک روز در نظر گرفته شد. این بازه زمانی از آنجا انتخاب شد که امکان دارد نمایه‌سازی تنها در فاصله چند ساعت پس از انتشار شماره جاری روزنامه صورت گیرد.

به‌منظور پرهیز از هر گونه ناهماهنگی، جستجو برای هر وبسایت و موتور کاوش در ساعات مشخصی از روز (بین ساعت ۶ تا ۹ صبح) تکرار می‌شد، زیرا ممکن بود یک نسخه حافظه پنهان که در ساعت ۸ صبح روزآمد نشده بود، در ساعات پایانی همان روز روزآمد شود. در این شرایط دو نرخ تازگی متفاوت به دست می‌آمد.

## ۵. تجزیه و تحلیل یافته‌ها

### ۱-۵. سرعت روزآمدسازی هر یک از موتورهای کاوش

به منظور تعیین سرعت روزآمدسازی هر موتور کاوش، میانگین نرخ تازگی ۲۵ وبسایت مورد بررسی در بازه زمانی مورد نظر به دست آمد. نتایج گردآمده در جدول ۱ نشان می‌دهد که گوگل پس از روزآمدشدن سایت مورد نظر، به‌طور میانگین در زمان کوتاه‌تری (۰/۶۸ روز) به روزآمدسازی نمایه یا نسخه پنهان خود دست می‌زند. به عبارت دیگر، گوگل در میان سایر موتورهای کاوش مورد بررسی دارای بیشترین سرعت روزآمدسازی است. یاهو (با میانگین ۳/۵ روز) و بینگ (با میانگین ۳/۷ روز) با فاصله زیاد پس از موتور کاوش گوگل قرار دارند. همچنین، گوگل تنها موتوری است که کمینه سرعت روزآمدسازی صفر را تجربه کرده است. به عبارت دیگر، این موتور کاوش، برخی از وبسایت‌های مورد بررسی را در همان تاریخ روزآمدسازی، نمایه‌سازی کرده است. همچنین، بیشینه فاصله زمانی بین انتشار روزنامه و نمایه‌شدن در وب برای گوگل (۳/۹ روز) از دو موتور دیگر کمتر است؛ بدین معنا که روزنامه‌های مورد بررسی حداکثر در حدود ۴ روز بعد از انتشار در گوگل نمایه شده‌اند. حال آنکه، این مقدار در موتورهای دیگر به حدود ۶ روز رسیده است.

جدول ۱. کمینه، بیشینه و میانگین سرعت روزآمدسازی موتورهای کاوش

موتور کاوش	سرعت روزآمدسازی	
	کمینه	بیشینه
گوگل	۰	۳/۹
ياهو	۰/۶۰	۵/۴۰
بینگ	۰/۶۷	۵/۷۷
کل	۰	۵/۷۷

### ۲-۵. روزآمدسازی بی‌درنگ موتورهای کاوش

وضعیت موتورهای کاوش به لحاظ فراوانی وبسایت‌های بی‌درنگ نمایه‌شده در جدول ۲ به تصویر کشیده شده است. چنانکه مشاهده می‌شود، ۲۳ عنوان (بالغ بر ۹۲ درصد

از کل) روزنامه‌های مورد بررسی بی‌درنگ به پایگاه نمایه موتورهای کاوش وارد شده‌اند. گوگل از دو موتور کاوش دیگر به لحاظ روزآمدسازی بی‌درنگ پیش‌تر است، به نحوی که بیش از ۹۰ درصد از وبسایت‌هایی که در فاصله زمانی یک روز یا کمتر نمایه شده‌اند، به این موتور تعلق دارند. این وبسایت‌ها، بیشینه وبسایت‌های مورد بررسی (۲۱ عنوان از ۲۵ عنوان بالغ بر ۸۸ درصد از کل روزنامه‌های فارسی) را تشکیل می‌دهند. دو موتور کاوش دیگر با نمایه‌سازی بی‌درنگ کمتر از ۵ درصد از وبسایت‌ها، با فاصله بسیار زیاد، از گوگل عقب افتاده‌اند.

### جدول ۲. وضعیت موتورهای کاوش به لحاظ نمایه‌سازی بی‌درنگ

موتور	وبسایت‌های بی‌درنگ روزآمد شده	
	فراوانی	درصد
گوگل	۲۱	۹۱/۳۰
ياهو	۱	۴/۳۵
بینگ	۱	۴/۳۵
جمع	۲۳	۱۰۰

### ۳-۵. مقایسه موتورهای کاوش به لحاظ سرعت روزآمدسازی روزنامه‌های فارسی

به منظور مقایسه میانگین سرعت روزآمدسازی وبسایت‌های روزنامه‌های فارسی در موتورهای کاوش از آزمون تحلیل واریانس استفاده شد. نتایج به دست آمده در جدول ۳ نشان داده شده است. همان گونه که از این جدول برمی‌آید، بین سه موتور کاوش به لحاظ میانگین سرعت روزآمدسازی وبسایت‌های روزنامه‌های فارسی اختلاف معناداری وجود دارد ( $F=۸۶/۷۶۱$ ،  $P=۰/۰۰۰$ ). به منظور شناسایی موتور کاوشی که اختلاف معنادار را ایجاد کرده است، از آزمون تعقیبی شفه استفاده شد. نتایج آزمون نشان داد که اختلاف میانگین سرعت روزآمدسازی موتورهای کاوش به گوگل باز می‌گردد، به نحوی که سرعت روزآمدسازی موتور کاوش گوگل با میانگین  $۰/۶۸$  روز با دو موتور کاوش دیگر، یعنی یاهو (با میانگین  $۳/۵۰$  روز) و بینگ (با میانگین  $۳/۷۰$  روز) فاصله بسیار زیادی دارد ( $P<۰/۰۵$ ). همان گونه که از مقدار میانگین دو موتور کاوش دیگر انتظار می‌رود، این دو به

لحاظ میانگین سرعت روزآمدسازی وبسایت‌های روزنامه‌های فارسی با یکدیگر اختلاف معناداری ندارند (جدول ۴).

جدول ۳. تحلیل واریانس برای سنجش معناداری اختلاف سرعت روزآمدسازی موتورهای کاوش

سطح معناداری	مقدار F	میانگین مربعات	درجه آزادی	مجموع مجددورات	
۰/۰۰۰	۸۶/۷۶۱	۷۱/۳۰۳	۲	۱۴۲/۶۰۷	بین گروهی
		۰/۸۲۲	۷۲	۵۹/۱۷۲	درون گروهی
			۷۴	۲۰۱/۷۷۹	مجموع

جدول ۴. موتورهای کاوش دارای میانگین سرعت روزآمدسازی متفاوت معنی دار در آزمون تعقیبی شفه

موتور کاوش (الف)	موتور کاوش (ب)	تفاوت میانگین‌ها	خطای استاندارد	سطح معناداری
گوگل	ياهو	-۲/۸۲	۰/۲۶	۰/۰۰۰
	بينگ	-۳/۰۲	۰/۲۶	۰/۰۰۰

#### ۵-۴. مقایسه سرعت روزآمدسازی وبسایت‌های روزنامه‌های فارسی

به منظور مقایسه میانگین سرعت روزآمدسازی وبسایت‌های روزنامه‌های فارسی در هر موتور کاوش از آزمون تحلیل واریانس استفاده شد. نتایج، در جدول ۵ به تصویر کشیده شده است. همان‌گونه که مشاهده می‌شود، در موتور کاوش گوگل بین وبسایت‌های روزنامه‌های فارسی به لحاظ میانگین سرعت روزآمدسازی اختلاف معناداری وجود دارد ( $P=۰/۰۰۰$  و  $F=۱۷/۶۹۰$ ). به منظور شناسایی وبسایت روزنامه‌هایی که این اختلاف معنادار را باعث شده است، از آزمون تعقیبی شفه استفاده شد. نتایج نشان داد که وبسایت روزنامه عصرآزادی (با میانگین نرخ تازگی ۳/۹ روز) با تمام وبسایت‌های مورد بررسی فاصله زیادی دارد ( $P<۰/۰۵$ ) و به این ترتیب، از کندترین سرعت روزآمدسازی برخوردار است. همچنین، میانگین نرخ تازگی روزنامه تفاهم (با میانگین ۱/۸ روز) با وبسایت روزنامه‌های کیهان (۰/۰۳ روز) و ابتکار، خبر و ایران (هر

کدام با میانگین نرخ تازگی صفر روز) به‌طور معناداری متفاوت است ( $P < 0/05$ ). دیگر وب‌سایت‌های روزنامه‌های مورد بررسی به‌لحاظ میانگین سرعت روزآمدسازی اختلاف معناداری با یکدیگر ندارند (جدول ۶).

جدول ۵. تحلیل واریانس برای مقایسه سرعت روزآمدسازی روزنامه‌های فارسی

موتور کاوش	مجموع مجذورات	درجه آزادی	میانگین مربعات	مقدار F	سطح معناداری
گوگل	بین گروهی	۲۴	۱۹/۴۷۸	۱۷/۶۹۰	۰/۰۰۰
	درون گروهی	۷۲۵	۱/۱۰۱		
	مجموع	۷۴۹			
ياهو	بین گروهی	۲۴	۲۶/۰۸۲	۸/۰۳۵	۰/۰۰۰
	درون گروهی	۷۲۵	۳/۲۴۶		
	مجموع	۷۴۹			
بینگ	بین گروهی	۲۴	۲۸/۴۰۵	۶/۸۴۱	۰/۰۰۰
	درون گروهی	۷۲۵	۴/۱۵۲		
	مجموع	۷۴۹			

جدول ۶. روزنامه‌های دارای میانگین سرعت روزآمدسازی متفاوت معنادار در موتور کاوش گوگل در آزمون تعقیبی شفه

روزنامه (الف)	روزنامه (ب)	تفاوت میانگین‌ها	خطای استاندارد	سطح معناداری
عصر آزادی	ابزار	۳/۱۷	۰/۲۷	۰/۰۰۰
	ابزار اقتصادی	۳/۳۰	۰/۲۷	۰/۰۰۰
	ابزار ورزشی	۳/۲۷	۰/۲۷	۰/۰۰۰
	آفریش	۳/۲۰	۰/۲۷	۰/۰۰۰
	افکار	۳/۴۷	۰/۲۷	۰/۰۰۰
	عصر اقتصادی	۳/۶۰	۰/۲۷	۰/۰۰۰

روزنامه (الف)	روزنامه (ب)	تفاوت میانگین‌ها	خطای استاندارد	سطح معناداری
	دنیای اقتصاد	۳/۳۷	۰/۲۷	۰/۰۰۰
	ابتکار	۳/۹۳	۰/۲۷	۰/۰۰۰
	اعتدال	۳/۶۷	۰/۲۷	۰/۰۰۰
	هدف و اقتصاد	۳/۷۰	۰/۲۷	۰/۰۰۰
	حمایت	۳/۷۳	۰/۲۷	۰/۰۰۰
	ایران	۳/۹۳	۰/۲۷	۰/۰۰۰
	جام جم	۳/۳۷	۰/۲۷	۰/۰۰۰
	جهان صنعت	۲/۴۰	۰/۲۷	۰/۰۰۰
	چوان	۳/۵۷	۰/۲۷	۰/۰۰۰
	کیهان	۳/۹۰	۰/۲۷	۰/۰۰۰
	خبیر	۳/۹۳	۰/۲۷	۰/۰۰۰
	مردم‌سالاری	۳/۴۷	۰/۲۷	۰/۰۰۰
	نسل فردا	۳/۴۰	۰/۲۷	۰/۰۰۰
	قدس	۳/۴۷	۰/۲۷	۰/۰۰۰
	رسالت	۳/۳۳	۰/۲۷	۰/۰۰۰
	سیاست روز	۲/۹۰	۰/۲۷	۰/۰۰۰
	تفاهم	۲/۱۳	۰/۲۷	۰/۰۰۰
	وطن امروز	۳/۰۰	۰/۲۷	۰/۰۰۰
تفاهم	ابتکار	۱/۸۰	۰/۲۷	۰/۰۰۹
	ایران	۱/۸۰	۰/۲۷	۰/۰۰۹
	کیهان	۱/۷۷	۰/۲۷	۰/۰۰۱
	خبیر	۱/۸۰	۰/۲۷	۰/۰۰۹

طبق نتایج جدول ۵، در یاهو نیز بین وب‌سایت‌های روزنامه‌های فارسی به‌لحاظ میانگین سرعت روزآمدسازی اختلاف معناداری وجود دارد ( $F= ۸/۰۳۵$  و  $P=۰/۰۰۰$ ).

نتایج آزمون تعقیبی شفه برای شناسایی وبسایت روزنامه‌ای که اختلاف معنادار را ایجاد کرده، نشان داد که وبسایت روزنامه‌ی *نسل فردا* (با میانگین نرخ تازگی ۰/۶ روز) با سرعت بیشتری نسبت به دیگر وبسایت‌های مورد بررسی روزآمد می‌شود ( $P < ۰/۰۵$ ). البته این اختلاف با روزنامه‌های *ایران*، *ایران ورزشی*، *ایران اقتصادی*، *افکار*، *حمایت*، *ایران*، *جام جم*، *جهان صنعت*، *جوان*، *خبر*، *قدس* و *تفاهم* معنادار نیست. دیگر وبسایت‌های روزنامه‌های مورد بررسی به لحاظ میانگین سرعت روزآمدسازی اختلاف معناداری با یکدیگر ندارند (جدول ۷).

جدول ۷. روزنامه‌های دارای میانگین سرعت روزآمدسازی متفاوت معنادار در موتور کاوش یا هو در آزمون تعقیبی شفه

روزنامه (الف)	روزنامه (ب)	تفاوت میانگین‌ها	خطای استاندارد	سطح معناداری
نسل فردا	آفرینش	-۴/۲۷	۰/۴۷	۰/۰۰۰
	عصر آزادی	-۳/۷۷	۰/۴۷	۰/۰۰۰
	عصر اقتصادی	-۴/۷۳	۰/۴۷	۰/۰۰۰
	دنیای اقتصاد	-۲/۹۳	۰/۴۷	۰/۰۲۵
	ابتکار	-۲/۸۳	۰/۴۷	۰/۰۴۷
	اعتدال	-۳/۰۰	۰/۴۷	۰/۰۱۶
	هدف و اقتصاد	-۳/۶۳	۰/۴۷	۰/۰۰۰
	کیهان	-۲/۸۳	۰/۴۷	۰/۰۴۷
	مردم‌سالاری	-۳/۱۷	۰/۴۷	۰/۰۰۵
	رسالت	-۳/۰۳	۰/۴۷	۰/۰۱۳
	سیاست روز	-۳/۰۳	۰/۴۷	۰/۰۱۳
	وطن امروز	-۴/۸۰	۰/۴۷	۰/۰۰۰

باز هم طبق نتایج حاصل در جدول ۵، در موتور کاوش بینگ نیز بین وبسایت‌های روزنامه‌های فارسی به لحاظ میانگین سرعت روزآمدسازی اختلاف معناداری وجود دارد ( $F = ۶/۸۴۱$  و  $P = ۰/۰۰۰$ ). نتایج آزمون تعقیبی شفه نشان داد که این اختلاف به سرعت

روزآمدسازی وبسایت روزنامه نسل فردا (با میانگین نرخ تازگی ۰/۶۶) باز می‌گردد که با وبسایت‌های آفرینش، افکار، عصر آزادی، عصر اقتصادی، هدف و اقتصاد، مردم‌سالاری و وطن امروز اختلاف معناداری دارد ( $P < 0/05$ ). دیگر وبسایت‌های روزنامه‌های مورد بررسی به لحاظ میانگین سرعت روزآمدسازی اختلاف معناداری با یکدیگر ندارند (جدول ۸).

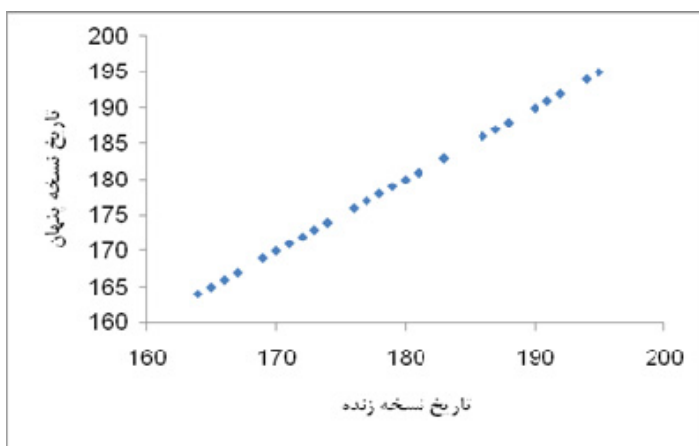
جدول ۸. روزنامه‌های دارای میانگین سرعت روزآمدسازی متفاوت معنادار در موتور کاوش بینگ در آزمون تعقیبی شفه

روزنامه (الف)	روزنامه (ب)	تفاوت میانگین‌ها	خطای استاندارد	سطح معناداری
نسل فردا	آفرینش	-۴/۸۷	۰/۵۲	۰/۰۰۰
	افکار	-۳/۲۳	۰/۵۲	۰/۰۴
	عصر آزادی	-۳/۸۰	۰/۵۲	۰/۰۰۱
	عصر اقتصادی	-۵/۱۰	۰/۵۲	۰/۰۰۰
	هدف و اقتصاد	-۳/۵۰	۰/۵۲	۰/۰۰۸
	مردم‌سالاری	-۳/۴۷	۰/۵۲	۰/۰۱
	وطن امروز	-۴/۸۰	۰/۵۲	۰/۰۰۰

#### ۵-۵. الگوی سرعت روزآمدسازی موتورهای کاوش

به منظور شناسایی الگوی حاکم بر سرعت روزآمدسازی روزنامه‌ها، آهنگ تغییرات تاریخ نسخه زنده وبی نسبت به تاریخ نسخه پنهان آنها در موتور کاوش بررسی شد. بدین منظور، نخست همبستگی میان تاریخ روزآمدسازی صفحه وب زنده روزنامه و تاریخ نسخه پنهان آن در هر موتور به کمک آزمون همبستگی پیرسون بررسی شد. آشکار است هر چه تاریخ صفحه وب زنده به تاریخ پنهان نزدیک‌تر باشد، روند بهتری بر روزآمدسازی حاکم است؛ به نحوی که در وضعیت آرمانی مقدار ضریب همبستگی یک یا بسیار نزدیک به یک خواهد بود. هر چه مقدار همبستگی از یک کمتر باشد، بدین معنی است که تاریخ نسخه پنهان نمایه موتور کاوش همگام با پیشروی تاریخ روزآمدسازی صفحه وب به پیش‌رفته و ممکن است چند روز ثابت مانده باشد. به عبارتی دیگر، با روزآمدسازی صفحه

وب زنده، نسخه پنهان تا چند روز روزآمد نمی‌شود. نتایج آزمون‌ها برای همه روزنامه‌ها معنادار بودن رابطه بین تاریخ نسخه زنده و نسخه پنهان را تأیید کرد ( $P=0/000$ ). در گام بعد، به کمک رگرسیون خطی قوت پیش‌بینی تاریخ نسخه پنهان بر اساس نسخه زنده مورد بررسی قرار گرفت. ضریب تعیین به دست آمده برای هر روزنامه در هر موتور کاوش در جدول ۹ آمده است. نظم جدول بر اساس الفبای عنوان روزنامه‌هاست. همان‌گونه که مشاهده می‌شود، ضریب تعیین به دست آمده از تحلیل رگرسیون خطی میان تاریخ صفحه وب زنده و تاریخ نسخه پنهان در گوگل برای روزنامه‌های ابتکار، ایران، خبر و کیهان برابر با یک است. این بدان معناست که نسخه پنهان این روزنامه‌ها در گوگل درست در همان تاریخ روزآمدسازی صفحه وب، روزآمد می‌شود. همچنین، در دیگر وب‌سایت‌های مورد بررسی نیز ضریب تعیین به یک بسیار نزدیک است؛ به نحوی که می‌توان ادعا کرد که در همه روزنامه‌ها، ۹۵ تا ۹۹ درصد از تغییرات نسخه‌های پنهان بر اساس تاریخ نسخه زنده صفحات وب روزنامه‌ها قابل پیش‌بینی است. این وضعیت کم‌وبیش در دو موتور کاوش دیگر نیز به چشم می‌خورد، با این تفاوت که هیچ‌یک از دو موتور کاوش یاهو و بینگ، ضریب تعیین برابر با یک نداشته‌اند و ضریب تعیین برخی روزنامه‌ها، مانند آفرینش یا هدف اقتصاد به ترتیب تا ۷۰ و ۸۱ درصد در یاهو و ۸۷ و ۸۲ درصد در بینگ کاهش یافته است.



نمودار ۱. الگوی روزآمدسازی نسخه پنهان روزنامه ابتکار در گوگل

نمودار ۱، الگوی روزآمدسازی نسخه پنهان روزنامه/تکرار در گوگل را نمایش می‌دهد. همان‌گونه که مشاهده می‌شود، آهنگ تغییرات تاریخ نسخه پنهان این روزنامه در گوگل کاملاً با تغییرات تاریخ نسخه زنده آن هماهنگ است. فاصله‌هایی که بین نقاط داده‌ای دیده می‌شود، به روزهای تعطیل رسمی اشاره دارد که روزنامه در آن روزها منتشر نشده است.

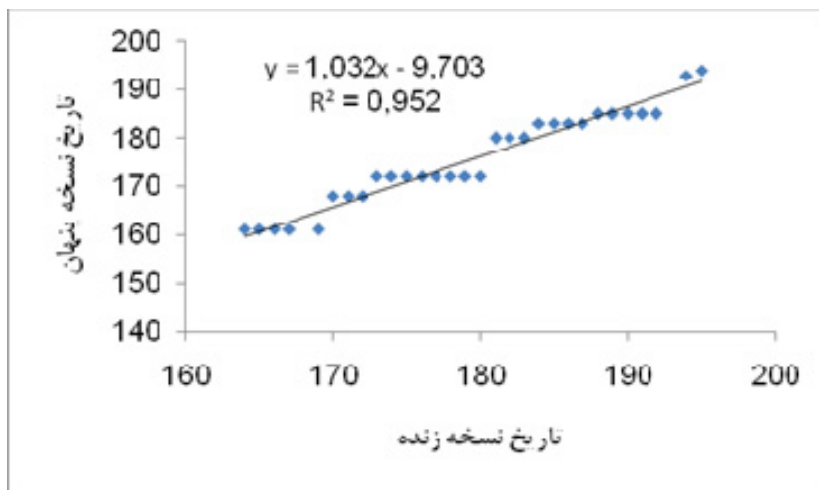
روزنامه عصر آزادی، کمترین مقدار ضریب تعیین را در گوگل به خود اختصاص داده است (۰/۹۵۲). با این حال، بررسی ضرایب تعیین به دست آمده برای این روزنامه (جدول ۹) نشان می‌دهد که سرعت نمایه‌سازی این روزنامه در یاهو و بینگ اندکی بهتر از گوگل است. این وضعیت درباره روزنامه دنیای اقتصاد نیز به چشم می‌خورد. همچنین، الگوی روزآمدسازی دو روزنامه رسالت و جهان صنعت در هر سه موتور کاوش مشابه است. گذشته از این معدود روزنامه، همه روزنامه‌های دیگر در گوگل بر پایه الگویی منظم‌تر نمایه می‌شوند، به نحوی که ضریب تعیین به دست آمده برای آنها در گوگل بالاتر از دو موتور کاوش دیگر است.

جدول ۹. قدرت پیش‌بینی روزآمدی روزنامه‌ها در موتورهای کاوش بر پایه تاریخ انتشار آنها

ردیف روزنامه	ضریب تعیین تعدیل شده			ردیف روزنامه	ضریب تعیین تعدیل شده		
	گوگل	ياهو	بینگ		گوگل	ياهو	بینگ
۱ آفرینش	۰/۹۷۹	۰/۷۰۸	۰/۸۷۷	۱۴ خبر	۱	۰/۹۸۸	۰/۹۷۷
۲ ابتکار	۱	۰/۹۶۰	۰/۹۷۲	۱۵ دنیای اقتصاد	۰/۹۸۱	۰/۹۹۰	۰/۹۸۷
۳ ابرار	۰/۹۸۴	۰/۹۶۱	۰/۹۶۱	۱۶ رسالت	۰/۹۸۳	۰/۹۸۳	۰/۹۸۵
۴ ابرار اقتصادی	۰/۹۸۵	۰/۹۶۱	۰/۹۶۱	۱۷ سیاست روز	۰/۹۹۲	۰/۹۸۸	۰/۹۸۱
۵ ابرار ورزشی	۰/۹۹۱	۰/۹۶۱	۰/۹۶۱	۱۸ عصر اقتصادی	۰/۹۷۳	۰/۹۴۰	۰/۹۵۱
۶ اعتدال	۰/۹۹۴	۰/۹۴۵	۰/۹۵۲	۱۹ عصر آزادی	۰/۹۵۲	۰/۹۷۴	۰/۹۶۸
۷ افکار	۰/۹۹۶	۰/۹۸۸	۰/۸۴۵	۲۰ قدس	۰/۹۹۷	۰/۹۹۰	۰/۹۸۱
۸ ایران	۱	۰/۹۵۸	۰/۹۶۶	۲۱ کیهان	۱	۰/۹۶۹	۰/۹۷۰
۹ تفاهم	۰/۹۸۴	۰/۹۷۸	۰/۹۸۳	۲۲ مردم‌سالاری	۰/۹۸۳	۰/۹۵۶	۰/۹۴۱
۱۰ جام جم	۰/۹۹۲	۰/۹۸۲	۰/۹۶۵	۲۳ نسل فردا	۰/۹۹۰	۰/۹۸۴	۰/۹۹۱
۱۱ جهان صنعت	۰/۹۷۵	۰/۹۷۵	۰/۸۵۲	۲۴ هدف اقتصاد	۰/۹۹۱	۰/۸۱۱	۰/۸۲۱

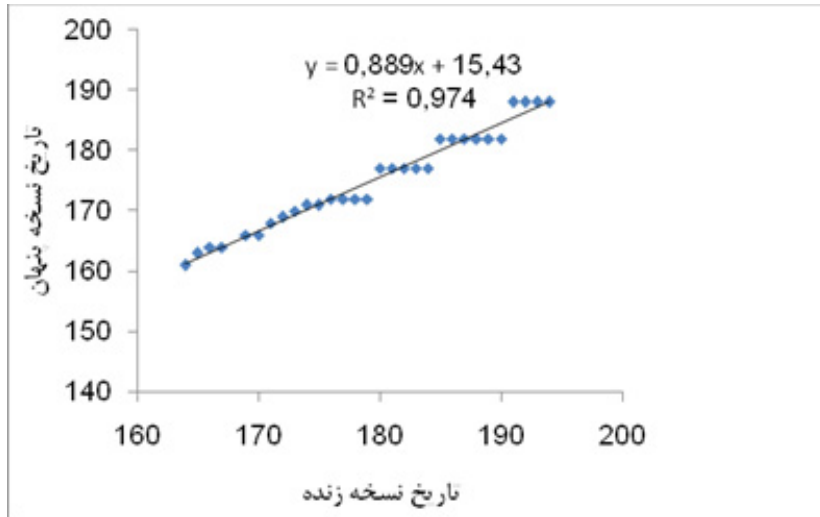
ردیف	روزنامه	ضرب تعیین تعدیل شده			ردیف	روزنامه	ضرب تعیین تعدیل شده		
		گوگل	یاهو	بینگ			گوگل	یاهو	بینگ
۱۲	جوان	۰/۹۹۶	۰/۹۷۷	۰/۹۸۵	۲۵	وطن امروز	۰/۹۷۴	۰/۹۲۵	۰/۹۳۲
۱۳	حمایت	۰/۹۹۲	۰/۹۵۶	۰/۹۶۲					

نگاهی به نمودارهای پراکنش<sup>۱</sup> تاریخ نسخه پنهان بر حسب تاریخ نسخه زنده برای روزنامه عصر آزادی در سه موتور کاوش یاهو، گوگل و بینگ (نمودارهای ۲ تا ۴) مؤید آن است که در هر سه موتور کاوش تاریخ نسخه‌های پنهان به مدت چند روز ثابت می‌ماند. الگوی روزآمدسازی این روزنامه در موتور کاوش گوگل در مقایسه با دو موتور دیگر اندکی بی‌نظمی بیشتری را نشان می‌دهد، به طوری که این روزنامه گاهی برای چندین روز پیاپی توسط روبات گوگل به فراموشی سپرده شده و روزآمد نشده است. همان‌گونه که از نمودار برمی‌آید، تأخیر در نمایه‌سازی این روزنامه در پایگاه گوگل سه، چهار، یا پنج روزه است و در یک مورد به ۸ روز نیز می‌رسد (نمودار ۲).

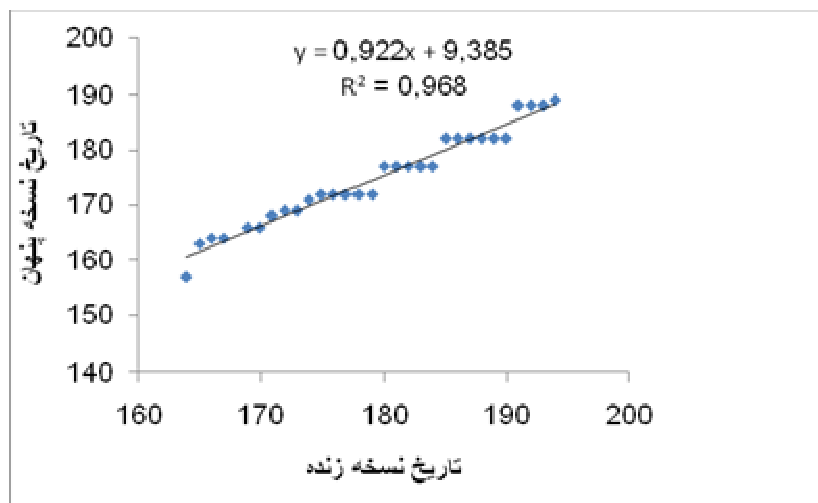


نمودار ۲. الگوی روزآمدسازی نسخه پنهان روزنامه عصر آزادی در گوگل

۱. به‌منظور درک بصری بهتر و همچنین امکان بررسی رابطه بین دو متغیر، تاریخ‌ها به تعداد روزهای گذشته از سال تبدیل شده است. برای نمونه، تاریخ ۸ مرداد ماه به  $۸ + (۳۱ \times ۵) = ۱۶۳$  روز



نمودار ۳. الگوی روزآمدسازی نسخه پنهان روزنامه عصر آزادی در یاهو



نمودار ۴. الگوی روزآمدسازی نسخه پنهان روزنامه عصر آزادی در بینگ

در دو موتور یاهو و بینگ (طبق نمودارهای ۳ و ۴)، در چند روز آغازین بازه زمانی مورد بررسی، تاریخ نسخه پنهان کم‌وبیش همگام با تاریخ نسخه زنده افزایش می‌یابد و از حدود اواسط بازه زمانی، تاریخ نسخه پنهان هر چند روز ثابت می‌ماند. حال آنکه، در موتور گوگل از همان آغاز تاریخ صفحه پنهان به مدت چندین روز ثابت می‌ماند. مدت زمان تأخیر در نمایه‌سازی این روزنامه در بینگ و یاهو تا شش روز است.

## ۶. نتیجه‌گیری

هیچ‌یک از ابزارهای جستجوی فعلی - از جمله موتورهای جستجو و راهنماها - قادر به پوشش دادن تمامی مطالب منتشر شده در اینترنت نیستند و شکاف میان تعداد صفحه‌های تولید شده و صفحه‌های نمایه‌سازی شده رو به افزایش است. طراحان صفحات با این هدف به انتشار در وب می‌پردازند که این صفحات توسط کاربران بالقوه وب پیگیری و مشاهده شوند. از آنجا که بیشتر کاربران از موتورهای جستجو بهره می‌گیرند، تنها زمانی می‌توان از بازایی روزآمد صفحه‌ها اطمینان حاصل کرد که صفحه‌ها در پایگاه‌های اطلاعاتی موتورهای کاوش نمایه شده باشند. هرچه سرعت نمایه‌سازی بالاتر باشد، دستیابی به خبر سریع‌تر است. مدت زمانی که میان آفرینش صفحه جدید یا بازنگری صفحه و شناسایی آن توسط موتور کاوش و ذخیره اطلاعات لازم در پایگاه اطلاعاتی موتور کاوش وجود دارد، پایه مفهوم سرعت روزآمدسازی است.

با آنکه در غالب حوزه‌های موضوعی و قلمروهای اطلاعاتی دستیابی به اطلاعات روزآمد حائز اهمیت است، در برخی قلمروها مانند خبررسانی، حساسیت و اهمیت اطلاعات به‌هنگام بسیار زیاد می‌باشد تا اندازه‌ای که شاید بتوان گفت حتی تأخیر بیش از یک روزه را بر نمی‌تابد. بنابراین، نمایه‌سازی بی‌درنگ می‌تواند خبررسانی به‌هنگام را تضمین کند.

به‌طور کلی، نتایج حاصل از پژوهش نشان داد که موتور کاوش گوگل به‌لحاظ سرعت روزآمدسازی روزنامه‌های فارسی رتبه بالاتری نسبت به موتورهای کاوش یاهو و بینگ دارد. این موتور، سریع‌ترین سرعت روزآمدسازی را داراست، به‌نحوی که تنها موتوری است که می‌تواند نسبت به روزآمدسازی نمایه خود در همان تاریخ روزآمدسازی صفحه اصلی اقدام کند. علاوه بر این، گوگل به‌طور کلی از الگوی روزآمدسازی

منظم تری پیروی می کند. اما موتور کاوش یا هو و بینگ علاوه بر سرعت کمتر نسبت به گوگل، الگوی روزآمدسازی نامنظم تری را به نمایش می گذارند. دو موتور کاوش یا هو و بینگ به لحاظ سرعت روزآمدسازی تفاوتی با یکدیگر ندارند. به نظر می رسد، موتور کاوش گوگل در دریافت اطلاعات روزآمد، کارآمدی بیشتری نسبت به دیگر رقبای خود دارد. این یافته همسو با یافته های لواندوسکی و همکاران است که نشان دادند گوگل در نمایه سازی وبسایت های آلمانی از دیگر موتورهای کاوش به روزتر است (Lewandowski et al. 2005). همچنین، لواندوسکی نیز شواهدی را درباره روزآمدی نمایه گوگل در مقایسه با یا هو ارائه کرده است (Lewandowski 2004). با آنکه مکیل و همکاران سرعت نمایه سازی اخبار خبرگزاری رویترز را در گوگل بسیار زیاد ارزیابی کرده اند، با این حال، آنان دریافتند که یا هو نیز از همان سرعت برخوردار است (Machill et al. 2005). احتمال می رود که درجه اهمیت خبرگزاری، دامنه پوشش خبری، گستره خبرپراکنی و همچنین زبان و مخاطبان اخبار در این تفاوت عملکرد یا هو نسبت به روزنامه های فارسی زبان در این میان مؤثر بوده باشد.

همچنین، در پژوهش حاضر صفحاتی مشاهده شد که چندین روز پیاپی در موتورهای کاوش گوناگون روزآمد نشدند. پرسشی که در این میان مطرح می شود، این است که سیاست موتورهای کاوش در به روزرسانی نمایه خود چیست و چه عاملی باعث می شود که نسخه پنهان صفحه ای در همان روز انتشار نمایه شده و برخی صفحات پس از چندین روز روزآمد شوند. به نظر می رسد که محبوبیت یک صفحه وب به روزآمدسازی سریع و بی درنگ آن منجر خواهد شد و صفحه ای را که محبوبیت کمتری دارد، ممکن است موتور کاوش با تأخیر نمایه کند. اما اینکه محبوبیت یک صفحه چگونه توسط موتور تعیین می شود، اهمیت بسیاری دارد. به نظر می رسد که محبوبیت یک صفحه وب از تعداد مشاهده کنندگان آن یا تعداد پیوندهایی که به صفحه وب مربوط داده می شود، اندازه گیری می شود (Lewandowski et al. 2005) پیشنهاد می گردد، پژوهشی به منظور تعیین رابطه میان محبوبیت صفحه وب و سرعت روزآمدسازی صفحات در موتورهای کاوش صورت گیرد تا اساس روزآمدسازی سریع تر برخی صفحات توسط موتورهای کاوش تعیین گردد.

به طور کلی، نتایج حاصل از این پژوهش نشان داد که موتور گوگل در دسترسی

به‌هنگام به محتوای روزنامه‌های فارسی بهتر از دو موتور دیگر عمل می‌کند. با این حال، با توجه به اینکه برخی روزنامه‌ها مانند عصر آزادی یا دنیای اقتصاد از الگوی روزآمدسازی منظمی در این پایگاه برخوردار نیستند، ضروری است خوانندگان این روزنامه جهت دسترسی به نسخه‌های روزآمدتر آن به یاهو یا بینگ نیز مراجعه کنند. همچنین، ضروری است متولیان روزنامه‌های فارسی، به‌ویژه این دو روزنامه، وبسایت‌های خود را ارزیابی کرده و بهینه‌سازی آنها بر پایه معیارهای نمایه‌شدن در موتورهای کاوش را مد نظر قرار دهند. مقایسه طراحی وبسایت روزنامه‌های فارسی با روزنامه/بتکار که به روزآمدسازی بی‌درنگ در گوگل دست یافته است، می‌تواند الهام‌بخش بهینه‌سازی وبسایت‌های دیگر روزنامه‌ها باشد.

### فهرست منابع

- پاپترو، اودیسپاس، پاپاستاورو، استاوروس، و ساماراس، جورج. ۱۳۸۴. نمایه‌سازی توزیع‌شده وب با استفاده از خزنده مهاجر. ترجمه رستم مظفری غربا. فصلنامه علوم و فناوری اطلاعات ۲۱ (۱): ۷۹-۸۶.
- داوودیان، فرشته. ۱۳۸۰. ابزارهای جستجو در اینترنت. تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی ۴۳: ۵۱-۶۳.
- شریف، عاطفه. ۱۳۹۱. بررسی تغییرات عملکرد دو موتور کاوش عمومی یاهو و گوگل از نظر پوشش کمی - زمانی نمایه‌سازی و توجه به عناصر ابرداده‌ای در رتبه‌بندی صفحه‌های وب. پژوهش‌نامه کتابداری و اطلاع‌رسانی ۲ (۱): ۱۷۵-۱۹۴.
- Lewandowski, D. 2004. Date-restricted queries in web search engines. *Online Information Review* 28 (6): 420-427.
- \_\_\_\_\_, H. Wahlig, and G. Meyer-Bautor. 2005. The freshness of web search engine databases. *Journal of Information Science* 32 (2): 133-150.
- Machill, M., D. Lewandowski, and S. Karzauninkat. 2005. Journalistische Aktualität im Internet. Ein Experiment mit den "News-Suchfunktionen" von Suchmaschinen. In: M. Machill and N. Schneider (eds.), *Suchmaschinen: Eine Herausforderung für die Medienpolitik*, (Vistas, Berlin).
- Notess, G. 2003. Search Engine Statistics: Freshness Showdown. Available at: <http://www.searchengineshowdown.com/stats/freshness.shtml> (accessed May 17, 2003).
- \_\_\_\_\_. 2002. Search Engine Statistics: Freshness Showdown. Available at: <http://www.searchengineshowdown.com/stats/0210freshness.shtml> (accessed October 20, 2002).
- \_\_\_\_\_. 2002. Search Engine Statistics: Freshness Showdown. Available at:

- <http://www.searchengineshowdown.com/stats/0204freshness.shtml> (accessed April 4, 2002).
- \_\_\_\_\_. 2002. Search Engine Statistics: Freshness Showdown. Available at:  
<http://www.searchengineshowdown.com/stats/0203freshness.shtml> (accessed March 7, 2002).
- \_\_\_\_\_. 2001. Search Engine Statistics: Freshness Showdown [Data from 13 August 2001].  
 Available at: <http://www.searchengineshowdown.com/stats/0108freshness.shtml> (accessed August 13, 2002).
- Ntoulas, A., J. Cho, and C. Olston. 2004. *What's New on the Web? The Evolution of the Web from a Search Engine Perspective*. In: Proceedings of the Thirteenth WWW Conference, New York, USA. [http://oak.cs.ucla.edu/~ntoulas/pubs/ntoulas\\_new.pdf](http://oak.cs.ucla.edu/~ntoulas/pubs/ntoulas_new.pdf) (accessed October 19, 2003).